A New Soft-Core Potential Function for Molecular Dynamics Applied to the Prediction of Protein Loop Conformations

K. TAPPURA,[1] M. LAHTELA–KAKKONEN,[2] O. TELEMAN[1,2,*]

[1]*VTT Biotechnology, P.O. Box 1500, FIN-02044 VTT, Finland*
[2]*CSC, Centre for Scientific Computing, P.O. Box 405, FIN-02101 Espoo, Finland*

Received 30 June 1999; accepted 2 November 1999

ABSTRACT: We have developed a new soft-core potential function for the conformational search of complex systems with molecular dynamics. The potential function was designed to maintain the main equilibrium properties of the original force field, which means that the soft-core potential gives physically realistic performance also without additional restraints, different from most of the previous soft-core potential functions. The performance of the method was demonstrated by applying it to the problem of finding native conformations for protein loops. Short loops from neocarzinostatin and parvalbumin were used as the first test cases. The use of the new soft-core potential function was shown to improve significantly the performance of molecular dynamics in the search of the native conformation of protein loops.
J Comput Chem 21: 388–397, 2000

Keywords: molecular dynamics; sampling of the conformational space; new soft-core potential; protein loop simulation with explicit water; prediction of protein loop structures

Correspondence to: K. Tappura; e-mail: kirsi.tappura@vtt.fi
*Present address: STFI, Swedish Pulp and Paper Research Institute, Box 5604, SE-11486 Stockholm, Sweden.
Contract/grant sponsor: European Commission (BIOTECH2) program; contract/grant number: BIO4-CT96-0166
Contract/grant sponsor: Academy of Finland

Introduction

The conformational space of a macromolecule, such as a protein, is very large and uneven, containing a huge amount of local energy minima surrounded by barriers. The conventional search methods, such as molecular dynamics (MD) and Monte Carlo (MC), sample only a small part of the conformational space due to their difficulties to overcome high energy barriers, or wide ones in the case of MC. The effective sampling of the conformational space is, thus, one of the main challenges of energy-based methods in the modeling of the three-dimensional structure of such molecules.

In fact, today homology modeling is usually the fastest practical way to build an approximate model for proteins for which the three-dimensional (3D) structure of related homologous proteins is available in databases to be used as templates. Homology modeling is based on the observation that the segments of similar conformations occur in proteins having related amino acid sequences. However, even for closely related proteins there are regions of unique conformations, which can vary widely both in sequence and size between the related proteins. Such regions are generally surface loops connecting the regular secondary structures. The fact that protein loops often form the site of biological activity and diversity makes the accurate modeling of loop structures not only challenging but also crucial.

There are two main strategies in the modeling of the 3D structure of proteins loops: the knowledge-based approach,[1–5] and the *ab initio* approach.[6–12] However, knowledge-based methods, in which the structural information is obtained from databases, have proven to be really powerful only for specific cases of loops, for example, for five of the six antibody loops.[3] In *ab initio* modeling no prior knowledge of the structures is used. Instead, a typical strategy is to generate a large number of possible conformations for the loop subject to imposed distance constraints to connect the loop to the rest of the protein and then screen or energy minimize to remove bad steric contacts, and finally, use energetic or other criteria to select the best conformation. However, even for relatively short loops the conformational space is very large, and a systematic search cannot cover the space properly to find the native structure. Stochastic search methods, such as molecular dynamics or Monte Carlo simulations, are needed to more effectively sample the conformational space, although there is still no guarantee that the minimum found is global and not local.

Several approaches have been developed to attempt to overcome the problem of multiple local minima. Perhaps the most frequently used approach is simulated annealing,[13] in which the kinetic energy of the system is increased to be able to cross over the barriers. Different methods for adding artificial degrees of freedom have also been developed to avoid the barriers existing in the three-dimensional space.[14–17] Other approaches, such as the diffusion equation method[18] and the application of various soft-core potential functions,[19–22] aim at smoothing the rough energy hypersurface. However, many of the soft-core potential functions presented in the literature[20, 21] replace the nonbonded interactions just by a soft repulsive term and omit the attractive part for simplicity. Although such an approach works well for structure refinement under distance restraints, it is not so useful in the case where no additional restraints are available. In addition to these simple repulsive soft-core potential functions, some more complicated functional forms have been developed for special purposes. Huber et al.[22] combined soft-core interaction functions, the diffusion equation method, and molecular dynamics. They selected the functional forms much from the point of view of the mathematical treatment of the diffusion equation method used for the additional smoothing of the energy surface. Beutler et al.[23] developed another soft-core potential function for free energy calculations to avoid singularities and numerical instabilities in thermodynamic integration. Because most of the soft-core potential functions were applied to NMR data refinement (or to other special purposes, which require special functional forms due to the mathematical complexity of the method), they do not concentrate on maintaining the properties of the original force field when no additional distance restraints are available, which, however, is crucial for *ab initio* loop modeling.

In this article, we present a novel soft-core potential energy function applied to modify a conventional molecular dynamics force field. The method was designed especially for the *ab initio* modeling of protein surface loops in their natural environment of water and the known structure of the protein framework, but can, of course, be applied to general conformational search problems with or without additional restraints.

The aim was to make the search problem of the conformational space more tractable by reducing the height of the barriers surrounding the energy minima without losing the main equilibrium properties of the original force field. Smoothing the

energy landscape should speed up and improve the convergence of the system to the (near) global minimum. Here, we concentrate on demonstrating the performance of the new soft-core potential for finding the native loop conformations compared to the known structures rather than on the important problem of identifying the right structure from the wrong ones in the real case when the correct structure is not known. Crambin was used as a training case to evaluate the physical limits of the parameters of the soft-core potential energy function. Six-residue loops of neocarzinostatin and parvalbumin were selected for the preliminary testing of the efficiency of the method.

Theory

Our aim was to smooth the potential energy surface to give atoms an additional freedom to occasionally quasi-penetrate each other. Equally important was to maintain the realistic physical interactions between atoms at interatomic distance at which the repulsion of the atom cores is no longer dominating (i.e., at the distance close to or above the sum of the van der Waals radii). The smoothing was applied to Lennard–Jones and Coulombic interaction functions by replacing the original potential by a function giving a finite value at zero interatomic distance. The other interactions were covered by the functional forms of the original GROMACS force field:[24] bond stretching, angle bending and improper dihedrals (i.e., the planarity of the groups) by harmonic potentials, and the rotations about single bonds (proper dihedrals) by sinusoidal energies.

To minimize the additional computational cost it was necessary to keep the functional form of the soft-core potential as simple as possible. In addition, the functional form had to allow the adjustment of the value of the soft-core potential energy at zero interatomic distance without disturbing the interactions at any distance above the sum of the van der Waals radii. After some experimentation with different functional forms, it was found that simple polynomial functions appear to fulfil all these requirements. To obtain the results presented in this article the following functional form was used for the soft-core potential energy.

$$V_{ij}^{\mathrm{soft}}(r_{ij}) = a_{ij} r_{ij}^{6} + b_{ij}, \qquad (1)$$

where r_{ij} is the distance between atoms i and j and a_{ij} and b_{ij} are the parameters of the soft-core potential interaction between atoms i and j.

In practice, the soft-core potential energy was included in the force field by smoothing the Lennard–Jones and Coulombic potential energy terms together. This enables easy adjustment of the height of the energy barrier. In the soft-core interaction function, the sum of the original Lennard–Jones and Coulombic energy terms, V_{ij}^{LJC}, was replaced by a function, $V_{ij}^{\mathrm{LJC\_soft}}$, depending on the interatomic distance, r_{ij}:

$$V_{ij}^{\mathrm{LJC\_soft}}(r_{ij}) = \begin{cases} V_{ij}^{\mathrm{soft}}(r_{ij}), & r_{ij} < r_{ij,\mathrm{switch}} \\ V_{ij}^{\mathrm{LJC}}(r_{ij}), & r_{ij} \geq r_{ij,\mathrm{switch}} \end{cases}, \qquad (2)$$

where

$$V_{ij}^{\mathrm{LJC}}(r_{ij}) = 4\varepsilon_{ij}\left(\frac{\sigma_{ij}}{r_{ij}^{12}} - \frac{\sigma_{ij}}{r_{ij}^{6}}\right) + \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_i q_j}{r_{ij}}. \qquad (3)$$

q_i and q_j are the charges of the particles i and j, respectively, ε_{ij} and σ_{ij} are the parameters of the Lennard–Jones interaction,[25] and $\varepsilon = \varepsilon_0\varepsilon_r$ is the dielectric constant.

The function $V_{ij}^{\mathrm{LJC\_soft}}$ itself as well as its derivative were made continuous at $r_{ij} = r_{ij,\mathrm{switch}}$ (to be exact, all the derivatives should be continuous, but only the first one is really critical for the method to work). The selected functional form of V_{ij}^{soft} as such takes care of the fact that the derivatives of the function are zero at $r_{ij} = 0$. The value of $r_{ij,\mathrm{switch}}$ depends on the type of the interaction (or on the interacting atoms). When there is a minimum in the V_{ij}^{LJC} function, the barrier height, b_{ij} (i.e., the value of the function $V_{ij}^{\mathrm{LJC\_soft}}$ at $r_{ij} = 0$), is predefined by the user, and $r_{ij,\mathrm{switch}}$, as well as a_{ij}, are solved numerically to fulfill the continuity conditions. Thus, the parameters of the soft-core potential depend on each other and on the parameters of V_{ij}^{LJC}. The implementation guarantees that $r_{ij,\mathrm{switch}}$ is always smaller than r_{ij} at the minimum of V_{ij}^{LJC}. If there is no minimum in V_{ij}^{LJC}, $r_{ij,\mathrm{switch}}$ is predefined by the user to be equal to a desired fraction ($k_{\_\mathrm{vdW}}$) of the sum of the van der Waals radii[26] of the interacting atoms: $r_{ij,\mathrm{switch}} = k_{\_\mathrm{vdW}}(r_{i,\mathrm{vdW}} + r_{j,\mathrm{vdW}})$. In this case, the potential energy $V_{ij}^{\mathrm{LJC\_soft}}$ at $r_{ij} = 0$, b_{ij}, can only be adjusted by varying $k_{\_\mathrm{vdW}}$, which means that it is impossible to lower the barrier height below a certain limit without pushing $r_{ij,\mathrm{switch}}$ to such a high value that the physical reality of the calculations is lost. A too high $r_{ij,\mathrm{switch}}$ is also likely to cause computational instabilities. One solution could be to use a higher degree polynomial for the soft-core potential, which allows $V_{ij}^{\mathrm{LJC\_soft}}$ to follow the potential function of the original force field to

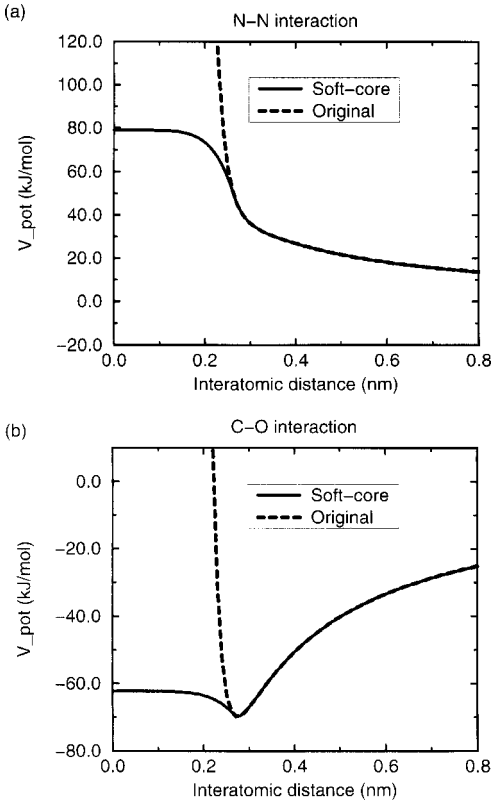

FIGURE 1. (a) The $V_{ij}^{\mathrm{LJC\_soft}}$ interaction energy of a pair of peptide nitrogen atoms, and (b) that of a peptide carbon and carbonyl oxygen atom compared to the sum of the Lennard–Jones and Coulombic energy terms of the original force field, V_{ij}^{LJC}.

smaller interatomic distances without forcing the barrier too high. However, preliminary tests with higher degree polynomials do not show significant improvement, which is probably due to the fact that, in any case, the energy at $r_{ij} = 0$ is very high when the Coulombic energy term is dominating the interaction of atoms having charges of equal sign. Figure 1 shows the $V_{ij}^{\mathrm{LJC\_soft}}$ interaction energy of a pair of peptide nitrogen atoms and that of a peptide carbon and carbonyl oxygen atom compared to the sum of the Lennard–Jones and Coulombic energy terms of the original force field, V_{ij}^{LJC}.

Methods

All the MD runs were performed with the GROMACS (version 1.6) molecular dynamics package,[24, 27] which was locally adjusted to include the option to use the soft-core potential energy function. The GROMACS force field (based on the GROMOS-87 force field parameters[28, 29] with some modifications[30, 31]) was used and will be referred to as the original molecular dynamics force field in this article. How to include the soft-core potential energy function and to obtain its parameters from those of the original force field was presented in detail above. Most of the MD runs were performed with the force field modified by the soft-core potential as described above, but some simulations were made with the original GROMACS force field for comparison. Unless stated otherwise, all the MD runs were performed using the following methods and parameters. Periodic boundary conditions were applied with a cubic box. The simulations were performed with a 2-fs time step, with all bond lengths constrained with the Lincs algorithm[32] in nonwater molecules and with SETTLE[33] in water molecules. The cutoff value for nonbonded interactions was 1 nm and the charge group-based pairlist was updated every 10 steps. The solvent (SPC water[34]) was included explicitly in the calculations to make the environment of the surface loops as realistic as possible. The temperature was controlled using weak coupling to a bath of 300 K with a time constant of 0.1 s.[35] The pressure was also controlled, using the similar procedure with a time constant of 0.5 ps. The initial velocities were randomly generated from a Maxwell distribution at 300 K. The computations were carried out mainly on Silicon Graphics Indigo2 (R10000) workstations and on a Cray T3E parallel machine, but partly also on a PC with a Pentium II (350 MHz) processor running Linux. The total CPU time required for the development of the method and for the preliminary testing presented in this article was approximately 30 weeks calculated in Indigo2 CPU time.

First, a crambin structure (1crn.pdb) was taken from the Protein Data Bank (PDB) to study and determine the physical limits of the parameters of the soft-core potential energy function. The PDB structure was used as the initial conformation for these MD test runs. After that, two X-ray crystal structures were selected from PDB to test the method for protein loop modeling. The resolution of the selected structures was below 0.16 nm with further requirements as follows: no crystal contacts were allowed in the loop area, loops had to be on the surface of proteins and have reasonably low mobility (low B-factor) and, finally, the structural features had to be good. The proteins were examined with the WHATIF program[36, 37] to check the quality of the structure compared with current reliable structures. Neocarzinostatin (1noa.pdb) and parvalbumin (5pal.pdb) were selected as the first

test proteins. The loops chosen for the studies were 1noa (25–30) and 5pal (33–38), with the numbers in parentheses indicating the residue sequential numbers. Several random conformations were created for each loop using the InsightII 97.0 program by MSI (Molecular Simulations Inc.), which randomly generates the dihedral angles of the loop backbone together with some adjustment to meet the loop-closure requirement, which means that the neighboring residues (anchor residues) on both sides of the loop were also somewhat shifted. The generated loops were energy minimized with 100–200 steps of steepest descent method using CHARMm[38] (Quanta 98) to get rid of high-energy conformations. Nonphysical conformations, such as loops through rings, loops still having residues of the opposite chirality or badly overlapping atoms after the energy minimization were rejected. Thus, the unsuccessful energy minimization served as a rejection criterion for bad conformations. The generated loop structures that were able to pass all the criteria were used as the starting conformations for the MD runs. The positions of the rest of the protein atoms were taken from the PDB file.

The loop of interest plus the anchor residues (the closest nonloop residues on both sides of the loop) were free to move under the force field during the MD runs, while the rest of the protein heavy atoms were restrained to their crystal positions by a harmonic potential with the force constant of 1000 kJ/mol/nm^2. From now on, the word "loop" is considered to include also the freely moving anchor residues, and as such, the MD runs are dealing with loops of eight residues. The soft-core potential, whenever used, was applied only to the protein atoms except in the crambin MD runs, where the solvent was also moving under the soft-core potential for the developmental purposes of the method. The protein was solvated in a cubic box such that the minimum distance between the protein and the edge of the box was 0.6 nm. The system was energy minimized using the force field containing the soft-core potential with relatively high barriers at $r_{ij} = 0$. After the minimization, 10 ps of MD was performed with harmonic constraints on the atomic coordinates of the protein to relax the water nearby. The final conformation was used as a starting structure of the real soft-core MD run. During the run the height of the soft-core energy barriers was changed after every 60 ps (or 30 ps in some experimental runs) so that the barriers were stepwise increased from a low value ($V_{ij}^{\mathrm{LJC\_soft}}(0) = 0.5kT$, $1kT$ or $3kT$ and $k_{\mathrm{\_vdW}} = 0.85$, where T is the absolute temperature, and k the Boltzmann constant) to a high one ($V_{ij}^{\mathrm{LJC\_soft}}(0) = 6kT$ and $k_{\mathrm{\_vdW}} = 0.6$) three times per run. The final conformation was energy minimized, and an additional MD run of 10–40 ps was performed with the original GROMACS force field. The final conformation of that run was again energy minimized using the original force field. The length of a typical MD simulation was 520 ps for a protein in water. Table I summarizes the soft-core MD run procedure with the parameters used.

For comparison and for the validation of the method the generated structures were also subject to MD under the original force field. Similar to the procedure described above for the soft-core MD, the solvating of the protein, energy minimization, and MD with the protein framework coordinates restrained were performed, but now with the original force field. The length of actual MD run was 400–500 ps with the original force field.

TABLE I.
The Run Procedure for MD with the Soft-Core Potential Energy Function.

	Barrier Height at $r_{ij} = 0$	$k_{\mathrm{\_vdW}}$	Duration (ps)
Soft MD	$3kT$	0.85	60
Soft MD	$6kT$	0.60	60
Soft MD	$1kT$	0.85	60
Soft MD	$3kT$	0.85	60
Soft MD	$6kT$	0.60	60
Soft MD	$0.5kT$	0.85	60
Soft MD	$3kT$	0.85	60
Soft MD	$6kT$	0.60	60
Energy minimization	infinite	—	max of 200 steps
Original MD	infinite	—	40
Energy minimization	infinite	—	max of 200 steps

Results and Discussion

When the functional form of the soft-core potential was fixed and implemented by adding the required new code to the GROMACS molecular dynamics program, physical limits were searched for the parameters of the soft-core potential function. For this purpose, a 46-amino acid protein, crambin (1crn), in water was used as a test system. The soft-core potential was applied to all the atoms, including water, to be able to monitor the behavior of the whole system with different soft-core parameters. The root-mean-square fluctuations (RMSF) were calculated as a function of the protein backbone atoms from a number of 120 ps MD runs performed with different soft-core parameters. The parameters were kept constant during each run. Some of those results are shown in Figure 2. Figure 2a shows how the flexibility of the molecule increases as the potential energy at $r_{ij} = 0$ is lowered. When the barrier height (at $r_{ij} = 0$) is varied only in those interactions where the Coulombic term is dominating (Fig. 2b), no change is observed in RMSF until $k_{\_vdW}$ approaches 0.95. Although the potential energy $V_{ij}^{LJC\_soft}$ at $r_{ij} = 0$ decreases significantly, as $k_{\_vdW}$ increases from 0.60 to 0.95, it is still very high, and the large RMSF values at $k_{\_vdW} = 0.95$ are due to the loss of physical reality of the interactions, rather than due to the soft-core atoms. This can be seen from the root-mean-square deviations (RMSD) studied with respect to the crambin (PDB) crystal structure (Fig. 2c and d), which also served as a starting conformation. The idea of these runs was to find the limiting values for the soft-core parameters that make the molecule as flexible as possible (RMSF as high as possible) but still reasonably well maintain the correct crystal structure during short runs. As Figure 2c and d indicate, it was found that $k_{\_vdW}$ must be below 0.95, preferentially below 0.90, and the barrier height above $1kT$ to avoid strange

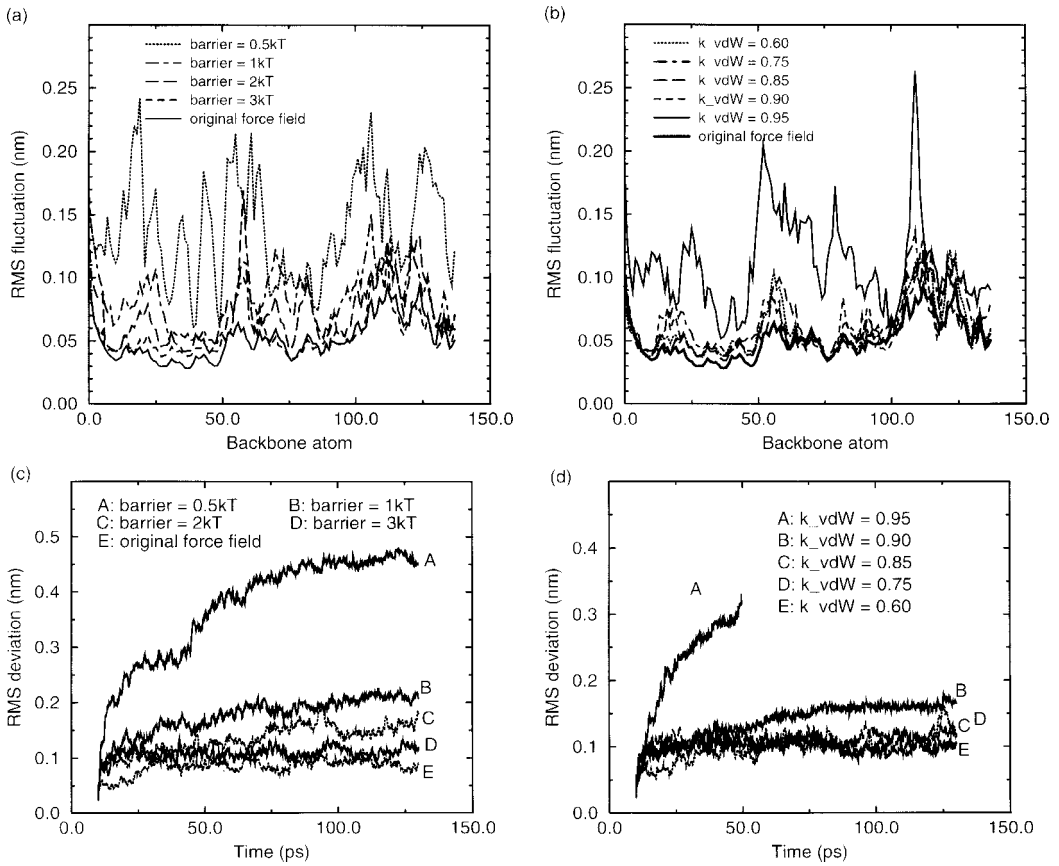

FIGURE 2. (a,b) The root-mean-square fluctuations (RMSF) of a crambin molecule as a function of the protein backbone atoms obtained from a number of 120 ps MD runs performed with different soft-core parameters and with the original force field. (c,d) The root-mean-square deviations (RMSD) for the same runs studied with respect to the crambin (PDB) crystal structure.

or unphysical structures. In the case, when only a loop is free to move and, thus, is subject to the restraints set by the rest of the protein, the system is not able to rapidly deviate from a realistic structure. Therefore, it is possible to enhance conformational changes in the loop by including periods with the barrier heights as low as $0.5kT$ and $1kT$ in the MD run (Table I). It is also worth mentioning that the implementation of the soft-core potential function made MD numerically very stable compared to the original force field. Even very unrealistic starting structures were not able to crash the run as long as the parameters of the soft-core potential function were in the allowed limits.

Figure 3 shows two examples of the several successful MD runs—one presenting the neocarzinostatin (24–31) loop, and the other the loop of parvalbumin (32–39). The PDB crystal structures were used as the reference conformations in the RMSD calculations without energy minimization for the corresponding data. The least-square fitting was done for the RMSD calculations by fitting the backbones of the protein frameworks (excluding the loop) on top of each other. As can be seen, the loop structures converge nicely near to the native conformations presented by the PDB structures. The RMS deviation is calculated both for the backbone atoms and for all the atoms of the loop. In both the cases RMSD decreases well below 0.15 nm. This kind of convergence to the correct structure can be considered exceptional in the modeling of the protein loop structures. In Figure 3b, the immediate effect of the very low barrier height ($0.5kT$) can also be seen as an increase in RMSD after the low RMS deviations were achieved for the first time. The regrowth of the barrier to the physical values (see Table I) is, however, able to draw the conformation back to the near native one. Nevertheless, this is not always the case. Figure 4 shows the initial and the final structure of the loop 1noa (24–31) compared with the PDB crystal structure for another successful soft-core MD run.

Figure 5 summarises the results of the two sets (1noa and 5pal) of MD runs starting from different initial loop structures. Again, the run procedure is given in Table I. All the initial loop conformations were subject to five MD runs starting with different velocities. Figure 5 shows the final RMSD of the 1noa and 5pal runs together with RMSD of the energy minimized starting structure. In addition to the final RMS deviation, the lowest RMSD is depicted for those runs where the final RMSD is significantly higher than the lowest value of the trajectory. It can be seen that the MD run with the soft-core potential energy function is almost always able to find a

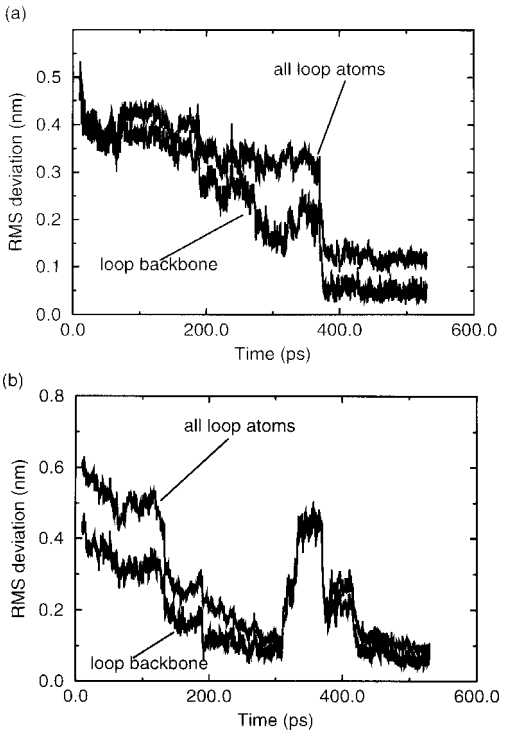

FIGURE 3. The RMS deviation of (a) the loop 1noa (24–31) and (b) the loop 5pal (32–39) as a function of time calculated with respect to the corresponding PDB crystal structure.

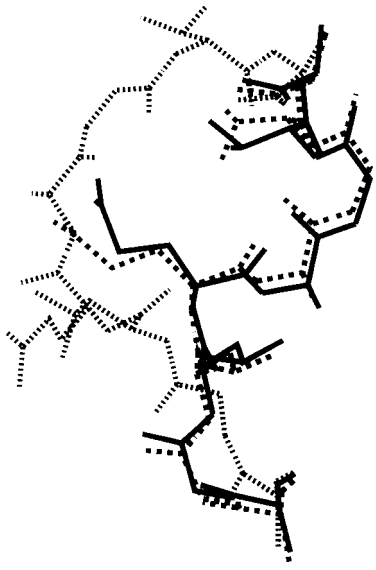

FIGURE 4. The initial (short dashed line) and final (dashed line) structures of the loop 1noa (24–31) compared with the PDB crystal structure (solid line) for a successful soft-core MD run.

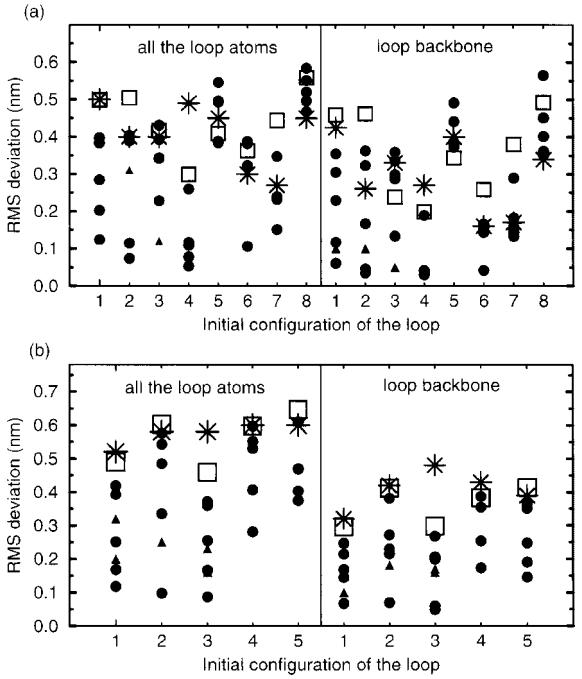

FIGURE 5. The final RMS deviations of the (a) 1noa and (b) 5pal runs (•) together with RMSD of the energy minimized starting structure (□) and that of the result of a typical original force field run (∗). The triangle shows the lowest RMSD for those runs where the final RMSD is significantly higher than the lowest value of the trajectory.

conformation with considerably lower RMSD compared to the energy minimized starting structure. The native conformation of the loop could be found for six of the eight starting conformations of 1noa and for three to five of the five starting conformations of 5pal, depending on whether the criterion is that the all-atom RMSD or the backbone RMSD is less than 0.2 nm. For comparison, the typical results of the MD runs performed with the original force field are also shown in the Figure 5. As can be seen from the figure, and was found from the tens of the MD trajectories, there is a very clear distinction between the behavior of the soft-core and original MD runs. Typically, MD with the original force field was not able to make significant changes in the loop conformation, in contrast to the soft-core MD runs (Fig. 3). However, some structures never converged to the native structure despite the large conformational changes they experienced during the soft-core MD run. A closer look at these conformations (loop conformations 5 and 8 in Fig. 5a) generated for 1noa show that, different from the other starting conformations, they both point the side chain of the same amino acid (leucine) in the opposite direction to that of the PDB structure, and that there is some steric hindrance preventing its rotation. Although the soft-core potential is often able to rotate the leucine side chain close to its crystal position, the structure is still so much deformed that the correct conformation cannot be found before the end of the run. This may be due to the fact that there are still relatively high energy barriers caused by the (original) dihedral angle energy terms preventing bond rotations in the soft-core force field.

Although thermodynamics states that the predominant conformation at any temperature is the one with the lowest free energy, the lowest potential energy is often considered to be an adequate criterion for the native conformation, as a first approximation. Therefore, the several different combinations of the potential energy terms were studied to see whether an energy-based criterion could also be found to distinguish the correct loop conformation when the native structure is not known. No clear correlation could be found between RMSD and any of the energy terms studied when the energy was calculated from the original GROMACS force field. The total intramolecular interaction energy of the protein or the total nonbonded interaction energy of the loop (i.e., intraloop interactions) appeared to give the best guesses in many cases, but failed totally in some others, showing lower energy for conformations far from the native one (Fig. 6). Even worse correlation was found between the RMS deviation and the interaction energies involving protein–solvent or loop–solvent interactions, probably due to the background noise caused by the thermal motion of the water molecules. When the interactions of the whole system, including all the solvent, were considered, the energy was completely dominated by the solvent–solvent interactions forming a high background from which the protein-related interactions could not be detected.

All these results showing weak or no correlation between RMSD and energy are in accordance with the observations of Daura et al.[39] for a heptapeptide and that of Zheng[40] for loop conformations. Two explanations are often suggested for this: a higher potential energy may be entropically more favored than the global minimum, and the native conformation may be a local minimum that is more easily reached during the folding process than the conformation at the global minimum. The poor correlation may also be related to the fact that the molecular dynamics force fields may have deficiencies in describing the relative energies of loop conformations in proteins. A demonstration of this can be seen in the recent report of Rapp et al. on the comparison of the AMBER* and AMBER94 force fields

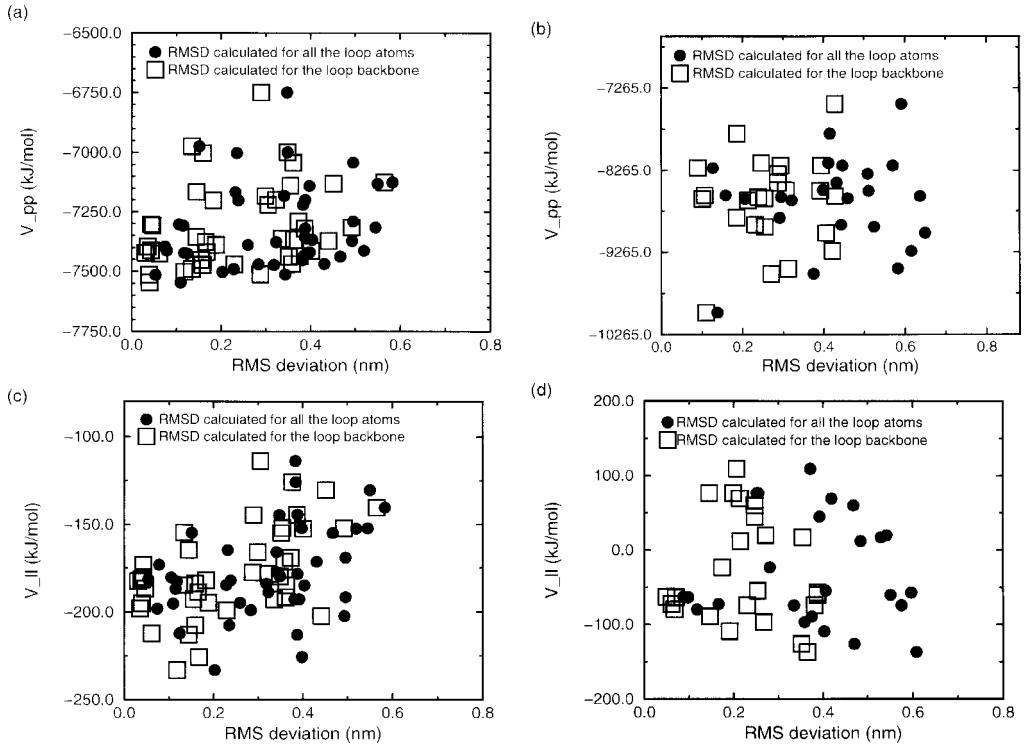

FIGURE 6. The total intramolecular interaction energy of (a) 1noa and (b) 5pal and the total nonbonded interaction energy of the loop (i.e., intraloop interactions) of (c) 1noa and (d) 5pal.

for the energetic ranking of loop conformations.[41] We should also like to emphasize the influence of the loop–solvent interactions: although the explicit water molecules make the calculations more realistic compared to the implicit models, and as such should give more accurate results, they are also the source of the high background noise in the system potential energy making the small energetic changes in the protein invisible, as mentioned above. Although disregarding the solvent interactions in the energy calculations seems an attractive alternative, which may even give reasonable results in some cases, correct energetic rankings cannot be expected from these simplified energies in general.

Conclusions

The new soft-core potential function was found to significantly improve the performance of molecular dynamics in the search of the native conformation of protein loops at least in the cases tested in this article. MD with the soft-core potential was able to find the native conformations in 100–500 ps for the six(eight)-amino acid loops of the two different proteins studied here. However, it should be noted that the soft-core MD, as well as the conventional MD, is a heuristic search method, which means that a single calculation does not guarantee good results. Therefore, several runs were performed with different starting conformations and starting velocities. The native conformation of the loop could be found for six of the eight starting conformations in the case of the first protein (1noa) and for three of the five starting conformations in the case of the second protein (5pal). In these cases, the conformation was considered native only when the all-atom RMS deviation from the PDB structure was less than 0.2 nm (the average RMSD was 0.12 nm for the loops with RMSD < 0.2 nm), which corresponds to the backbone RMSD of less than 0.15 nm (most cases less than 0.1 nm). If the backbone RMSD < 0.2 nm is used as a criterion, all the 5pal loop structures can be considered to be able to find the native conformation within five runs.

It was found that the quality of the initial loop conformation has an influence on the performance of the method. This means that the starting conformation should be more or less realistic so that structures with very high energy should not be used. Although in many cases the bad structures could be found and rejected by simple manual screening, the performance of the soft-core potential is believed

to improve, if another method, instead of the random generator combined with the rough manual screening, were used to generate intelligent starting structures. In addition, because only the Lennard–Jones and Coulombic energy terms were made soft, there are still other interactions forming high energy barriers in the force field. From the remaining energy terms, the contribution of the dihedral angle interaction appears to be most important for the conformational transitions. Therefore, the introduction of an adjustable-barrier dihedral angle potential would be a logical continuation to our work, together with further studies to find criteria to identify the near-native loop structures without prior knowledge.

Because the soft-core potential function was tested only with two six(eight)-residue loops, it is clear that the results presented in this article cannot be regarded as a statistical demonstration of the performance of the method. However, the results, even as such, are promising, and encourage further studies and testing of the method with a larger sample of different and longer loops.

Acknowledgments

The Cray T3E parallel computer was provided for the calculations by the CSC—Center for Scientific Computing, Finland.

References

1. Summers, N. L.; Karplus, M. Methods Enzymol 1991, 202, 156.
2. Reczko, M.; Martin, A. C. R.; Bohr, H.; Suhai, S. Protein Eng 1995, 8, 389.
3. Chothia, C.; Lesk, A. M.; Tramontano, A.; Levitt, M.; Smith–Gill, S. J.; Air, G.; Sheriff, S.; Padlan, E. A.; Davies, D.; Tulip, W. R.; Colman, P. M.; Spinelli, S.; Alzari, P. M; Poljak, R. J. Nature 1989, 342, 877.
4. Kwasigroch, J.-M.; Chomilier, J.; Mornon, J.-P. J Mol Biol 1996, 259, 855.
5. Donate, L. E.; Rufino, S. D.; Canard, L. H. J.; Blundel, T. I. Protein Sci 1996, 5, 2600.
6. Martin, A. C. R.; Cheetham, J. C.; Rees, A. R. Proc Natl Acad Sci 1989, 86, 9268.
7. Bruccoleri, R. E.; Haber, E.; Novotny, J. Nature 1988, 335, 564.
8. Dudek, M. J.; Scheraga, H. A. J Comput Chem 1990, 11, 121.
9. Zhang, H.; Lai, L.; Wang, L.; Han, Y.; Tang, Y. Biopolymers 1997, 41, 61.
10. Shenkin, P.; Yarmush, D. L. et al. Biopolymers 1987, 26, 2035.
11. Zheng, Q.; Kyle, D. J. Proteins 1996, 24, 209.
12. Collura, V.; Higo, J.; Garnier, J. Protein Sci 1993, 2, 1502.
13. Kirkpatrick, S.; Gelatt, Jr., C. D.; Vecchi, M. P. Science 1983, 220, 671.
14. Crippen, G. M. J Phys Chem 1987, 91, 6341.
15. van Schaik, R. C.; Berendsen, H. J. C.; Torda, A. E.; van Gunsteren, W. F. J Mol Biol 1993, 234, 751.
16. Crippen, G. M.; Havel, F. T. J Chem Inf Comput Sci 1990, 30, 222.
17. Purisima, E. O.; Scheraga, H. A. Proc Natl Acad Sci USA 1986, 83, 2782.
18. Piela, L.; Kostrowicki, J.; Scheraga, H. A. J Phys Chem 1989, 93, 3339.
19. Levitt, M. J Mol Biol 1983, 170, 723.
20. Nilges, M.; Clore, G. M.; Gronenborn, A. M. FEBS Lett 1988, 239, 129.
21. Ullner, M.; Selander, M.; Persson, E.; Stenflo, J.; Drakenberg, T.; Teleman, O. Biochemistry 1992, 31, 5974.
22. Huber, T.; Torda, A. E.; van Gunsteren, W. F. J Phys Chem A 1997, 101, 5926.
23. Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Chem Phys Lett 1994, 222, 529.
24. van der Spoel, D.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; van Drunen, R.; Berendsen, H. J. C. GROMACS User Manual version 1.5; Nijenbrgh 4: 9747 AG Groningen, The Netherlands. Inernet: http://rugmd0.chem.rug.nl/~gmx, 1997.
25. Allen, M. P.; Tildesley, D. J. Computer Simulations of Liquids; Oxford University Press: New York, 1987.
26. Bondi, A. J Phys Chem 1964, 68, 441.
27. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. Comput Phys Commun 1994, 91, 43.
28. van Gunsteren, W. F.; Daura, X.; Mark, A. E. In Encyclopedia of Computational Chemistry 2; John Wiley & Sons: New York, 1998; p. 1211.
29. van Gunsteren, W. F.; Berendsen, H. J. C. GROMOS87 manual; Biomos BV Nijenborgh 4: 9747 AG Groningen, The Netherlans, 1987.
30. van Buuren, A. R.; Marrink, S. J.; Berendsen, H. J. C. J Phys Chem 1993, 97, 9206.
31. van der Spoel, D.; van Buuren, A. R.; Tieleman, D. P.; Berendsen, H. J. C. J Biomol NMR 1996, 8, 229.
32. Hess, B.; Beker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. J Comp Chem 1997, 18, 1463.
33. Miyamoto, S.; Kollman, P. A. J Comput Chem 1992, 13, 952.
34. Berendsen, H. J. C.; Potsma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In Intermolecular Forces; D. Reidel Publishing Company: Dordrecht, 1981; p. 331.
35. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684.
36. Vriend, G. J Mol Graph 1990, 8, 52.
37. Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. Nature 1996, 381, 272.
38. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
39. Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. J Mol Biol 1998, 280, 925.
40. Zheng, Q.; Rosenfeld, R.; Vajda, S.; DeLisi, C. J Comput Chem 1993, 14, 556.
41. Rapp, C. S.; Friesner, R. A. Proteins 1999, 35, 173.